104526 + 139129

TEXTE POUR EXTENSIONS

ABREGE

La présente invention concerne un ordonnanceur, encore appelé discipline de service, pour un système comportant une pluralité de nœuds partageant une pluralité de ressources telles que des longueurs d'onde. L'ordonnanceur 2 selon l'invention ordonnance la transmission de données à partir d'une pluralité de files d'attente B_1, B_2 et B_3 d'un nœud source 1 vers une pluralité de nœuds de destination N_1, N_2 et N_3 via une pluralité de ports de sortie P_1, P_2, P_3 et P_4 dudit nœud source 1, chacun desdits ports de sortie P_1, P_2, P_3 et P_4 étant associés à une ressource OR_1, OR_2, OR_3, et OR_4, les données étant transmises via ladite ressource vers un nœud de destination N_1, N_2 et N_3, chacun desdits nœuds recevant des données de tout ou partie de ladite pluralité de ressources OR_1, OR_2, OR_3 et OR_4. Le dispositif ordonnanceur 2 est caractérisé en ce qu'il comporte une pluralité de serveurs S_1, S_2, S_3 et S_4, chacun desdits serveurs étant associé respectivement à une ressource de ladite pluralité de ressources OR_1, OR_2, OR_3 et OR_4 et chacun desdits serveurs comportant des moyens d'ordonnancement, lesdits moyens d'ordonnancement étant indépendants pour chacun desdits serveurs.

Figure à publier : Figure 1

Dispositif ordonnanceur pour un système à ressources partagées asymétriquement

La présente invention concerne un ordonnanceur, encore appelé discipline de service, pour un système comportant une pluralité de nœuds partageant une pluralité de ressources telles que des longueurs d'onde.

Un tel système est par exemple un réseau en anneau de paquets optiques du type DBORN (Dual Bus Optical Ring Network). L'architecture de l'anneau est organisée autour d'un concentrateur et est constituée d'une pluralité de nœuds tels que des multiplexeurs d'insertion/extraction de paquets optiques du type OPADM (Optical Packet Add/Drop Multiplexer), chaque nœud étant en communication avec le concentrateur. Ce réseau contient un bus d'écriture correspondant à une pluralité de longueurs d'onde dites montantes et un bus de lecture correspondant à une pluralité de longueurs d'onde descendantes. Les longeurs d'onde montantes et descendantes qui sont le plus souvent multiplexées sur une même fibre, sont utilisées et donc partagées par les nœuds du réseau pour émettre et recevoir des paquets vers et depuis le concentrateur. Une pluralité de nœuds partage donc une même ressource telle qu'une longueur d'onde pour recevoir des paquets envoyés par le concentrateur assimilable à un nœud source.

Cependant, pour tenir compte des spécificités de chacun de nœuds, tous les nœuds ne partagent pas nécessairement une même ressource. Ainsi, il se peut qu'une ressource soit partagée par une partie seulement des nœuds du réseau.

Chacun des nœuds ne partageant pas les mêmes ressources que les autres nœuds dans les mêmes proportions, on parle alors de ressources partagées asymétriquement.

Une des fonctions des réseaux concerne la discipline de service c'est à dire le fait de déterminer, parmi une pluralité de files d'attente ou buffers, quel est le paquet associé à une file d'attente particulière qui doit être envoyé sur un nœud. Cette détermination est réalisée par un dispositif appelé ordonnanceur.

La présente invention a pour objet un dispositif ordonnanceur, dit encore discipline de service, pour un système comportant une pluralité de nœuds partageant asymétriquement une pluralité de ressources telles que des longueurs d'onde.

La présente invention propose à cet effet un dispositif ordonnanceur pour ordonnancer la transmission de données à partir d'une pluralité de files d'attente d'un nœud source vers une pluralité de nœuds de destination via une pluralité de ports de sortie dudit nœud source, chacun desdits ports de sortie étant associés à une ressource, les données étant transmises via ladite ressource vers un nœud de destination, chacun desdits nœuds recevant des données de tout ou partie de ladite pluralité de ressources, ledit dispositif

ordonnanceur étant caractérisé en ce qu'il comporte une pluralité de serveurs, chacun desdits serveurs étant associé respectivement à une ressource de ladite pluralité de ressources et chacun desdits serveurs comportant des moyens d'ordonnancement, lesdits moyens d'ordonnancement étant indépendants pour chacun desdits serveurs.

Grâce à l'invention, chaque serveur fonctionne indépendamment des autres serveurs et peut prendre en compte les spécificités de la ressource à laquelle il est associé et notamment le fait qu'une ressource ne soit pas partagée de façon uniforme par tous les nœuds de destination, chaque nœud utilisant ladite ressource suivant un certain coefficient de pondération. Ce coefficient de pondération peut être nul si le nœud n'utilise pas ladite ressource. Il peut également être lui-même pondéré selon l'importance que prend ladite ressource pour le nœud de destination. Ainsi une ressource utilisée par un premier nœud et un deuxième nœud ne sera pas partagé de la même façon entre le premier nœud et le deuxième nœud si le premier nœud utilise davantage d'autres ressources que le deuxième nœud. Chaque serveur peut par exemple prendre en considération une double pondération : une première pondération donnant une information sur l'utilisation de la ressource par le nœud et traduisant l'asymétrie du système et une deuxième pondération donnant une information sur le ratio d'utilisation de la ressource par le nœud en fonction du trafic à destination dudit nœud par rapport au trafic total.

Selon un mode de réalisation, lesdits moyens d'ordonnancement comportent une pluralité d'étages correspondant respectivement à une pluralité d'ordonnancements selon des critéres distincts.

Selon un mode de réalisation, lesdits moyens d'ordonnancement comportent des moyens d'ordonnancement cyclique du type Round Robin.

Les moyens d'ordonnancement Round Robin parcourent séquentiellement et cycliquement des files d'attente de type FIFO (First In First Out) et servent la première file prête ou non vide. Si la file d'attente est vide, les moyens d'ordonnancement passent à la suivante. Certaines files peuvent être privilégiées en définissant un poids correspondant par exemple au nombre d'éléments ou paquets que peut prendre l'ordonnanceur en tête de la file d'attente ; on parle alors de Weighted Round Robin WRR.

Selon un autre mode de réalisation, lesdits moyens d'ordonnancement comportent des moyens d'ordonnancement WFR (Weighted Fair Queueing).

Cet algorithme donne un traitement prioritaire aux flux de faible volume et permet aux flux de volume important d'utiliser la place qui reste. Pour cela, il trie et regroupe les paquets par flux, puis met ceux-ci en file d'attente suivant le volume de trafic dans chaque flux.

Avantageusement, lesdits moyens d'ordonnancement sont dépendants d'un ensemble de pondérations statiques et/ou dynamiques.

Les pondérations statiques peuvent par exemple être issues de procédés classiques de partition ou d'allocation des ressources. Les pondérations dynamiques peuvent être calculées sur la base d'informations de contrôle de congestion. Une combinaison de ces deux types de pondérations peut également être envisagée.

Selon un mode de réalisation particulièrement avantageux, lesdits moyens d'ordonnancement sont dépendants d'un premier ensemble de pondérations, chacune desdites pondérations traduisant le pourcentage d'allocation de ladite ressource à chacun desdits nœuds de ladite pluralité de nœuds.

Ce type de pondération est obtenu par des procédés classiques de partition ou d'allocation des ressources.

De manière avantageuse, lesdits moyens d'ordonnancement sont dépendants d'un deuxième ensemble de pondérations, chacune desdites pondérations traduisant le poids relatif du trafic de chacun desdits nœuds par rapport au trafic total.

La présente invention a également pour objet un nœud comportant un dispositif ordonnanceur selon l'invention et incluant une pluralité de files d'attente pour l'émission de données vers une pluralité de nœuds de destination et une pluralité de ports de sortie.

La présente invention a en outre pour objet un système de transmission de données comportant au moins un nœud source selon l'invention, ledit système comportant :

- une pluralité de nœuds de destination,
- une pluralité de ressources.

D'autres caractéristiques et avantages de la présente invention apparaîtront dans la description suivante d'un mode de réalisation de l'invention, donné à titre illustratif et nullement limitatif.

- La figure 1 représente schématiquement un système de transmission incorporant un premier exemple de réalisation du dispositif ordonnanceur selon l'invention.
- La figure 2 représente schématiquement un système de transmission incorporant un second exemple de réalisation du dispositif ordonnanceur selon l'invention.
- La figure 3 illustre un arbitrage à trois niveaux.

La figure 1 représente schématiquement un système de transmission 10 tel qu'un réseau en anneau de paquets optiques. Cette représentation est limitée à la description de l'invention, ledit système pouvant comporter de nombreux autres éléments. Le système 10 comporte :

- un nœud source 1,
- trois nœuds de destination N_1, N_2 et N_3,

- quatre ressources OR_1, OR_2, OR_3 et OR_4.

Les ressources OR_1, OR_2, OR_3 et OR_4 sont par exemple des longueurs d'onde multiplexées sur une fibre optique selon une technologie DWDM (Dense Wavelength Division Multiplex).

Les nœuds N_1, N_2 et N_3 sont par exemple des multiplexeurs OPADM (Optical Packet Add/Drop Multiplexer).

Le nœud source 1 est par exemple concentrateur électronique tel qu'un commutateur Ethernet.

Le nœud source 1 comprend :

- trois files d'attentes ou buffers B_1, B_2 et B_3 permettant de stocker des paquets avant de les émettre respectivement vers les nœuds N_1, N_2 et N_3,
- un dispositif ordonnanceur 2 encore appelé discipline de service,
- quatre ports de sortie P_1, P_2, P_3 et P_4.permettant d'émettre les paquets de données respectivement sur les ressources OR_1, OR_2, OR_3 et OR_4.

Le dispositif ordonnanceur 2 comprend quatre serveurs S_1, S_2, S_3 et S_4 associés chacun respectivement aux ressources OR_1, OR_2, OR_3 et OR_4 et aux ports P_1, P_2, P_3 et P_4.

Chacun des quatre serveurs S_1, S_2, S_3 et S_4 détermine quel est le paquet associé à une file d'attente particulière qui doit être envoyé sur un nœud via la ressource associée au serveur.

Les ressources OR_1 et OR_2 sont partagées par les nœuds N_1 et N_2.

La ressource OR_3 est partagée par les nœuds N_2 et N_3.

La ressource OR_4 est partagée par les nœuds N_1 et N_3.

Les ressources ne sont donc pas partagées uniformément par les nœuds N_1, N_2 et N_3.

Ainsi, une même ressource utilisée par un premier nœud et un deuxième nœud ne peut pas être utilisée de la même façon par le premier nœud utilisant davantage d'autres ressources que le deuxième nœud.

Par exemple, le nœud N_1 utilise les ressources OR_1, OR_2 et OR_4 tandis que le nœud N_3 utilise uniquement les ressources OR_3 et OR_4. Le nœud N_1 peut donc utiliser trois ressources pendant que le nœud N_3 ne peut en utiliser que deux.

Le procédé d'allocation de ressources prend donc en compte cette allocation non uniformément répartie et attribue une pondération à chacun des nœuds correspondant au pourcentage d'allocation de ladite ressource à chacun desdits nœuds de ladite pluralité de nœuds. Cette pondération est notée, de façon générale, R_{ij} et correspond au ratio alloué au nœud N_i sur la ressource OR_j.

De plus, les nœuds de destination peuvent avoir des poids différents à cause de leurs trafics. Ainsi, en appelant T_i le trafic à destination du nœud N_i, chaque nœud peut être pondéré par un

coefficient W_i égal à (T_i / $\sum i$ T_i) où $\sum i$ T_i désigne la somme des trafics à destination de l'ensemble des nœuds.

Ainsi, chacun des serveurs se voit attribuer une série de pondérations, dites méta-pondérations, pour chacun des nœuds en prenant en considération à la fois l'asymétrie du partage des ressources et le trafic différent pour chacun des nœuds.

Ces méta-pondérations sont résumées dans le tableau 1 ci-dessous et correspond au produit de R_{ij} par W_i.

Serveurs / Nœuds	N1	N2	N3
s1	W1 x R11	W2 x R21	W3 x R31
s2	W1 x R12	W2 x R22	W3 x R32
s3	W1 x R13	W2 x R23	W3 x R33
s4	W1 x R14	W2 x R24	W3 x R34

Tableau 1

Chacun desdits serveurs utilise ces méta-pondérations et procède, de façon indépendante des autres serveurs, à un mécanisme d'ordonnancement du type Round Robin, WRR (Weighted Round Robin) ou WFR (Weighted Fair Queueing) afin de sélectionner la file d'attente et le ou les paquets à émettre. Les serveurs peuvent inclure des moyens logiciels, matériels ou une combinaison des deux.

Les pondérations telles qu'elles ont été décrites plus haut peuvent être mises à jour statiquement ou dynamiquement. Une mise

à jour dynamique permet une adaptation dynamique de l'ordonnancement en prenant en compte la variation de charge en fonction du temps et de la destination.

De plus, l'invention permet de préserver l'ordre des paquets en éliminant le besoin de mécanismes ou procédures complexes et coûteuses pour palier à un déséquencement et pour réorganiser les paquets. Pour assurer la préservation de l'ordre des paquets, il suffit que le service des paquets respecte l'ordre établi grâce à un accès parallèle paquet par paquet par les serveurs (et non par bloc).

L'invention a été décrite en relation avec un ensemble de pondérations traduisant le poids relatif du trafic de chacun des nœuds par rapport au trafic total mais d'autres ensembles de pondérations peuvent être utilisés traduisant d'autres paramètres ou caractéristiques de chacun des nœuds, tels que les types de service et/ou d'utilisateur. Les pondérations peuvent être appliquées sous la forme de méta-pondérations, comme cela est décrit ci-dessus, mais peuvent être appliquées aussi bien sous la forme de paramètres séparés à différents niveaux..

La **figure** 2 représente schématiquement un système de transmission incorporant un second exemple de réalisation du dispositif ordonnanceur selon l'invention, comportant une pluralité d'étages L1, L2, L3 correspondant respectivement à une pluralité d'ordonnancements selon des critéres distincts. Le réseau 10' est

analogue au réseau 10 décrit précédemment. Il diffère par le dispositif d'ordonnancement dans le nœud source 1', et il comprend :

- trois files d'attentes ou buffers B'_1, B'_2 et B'_3 permettant de stocker des paquets avant de les émettre respectivement vers les nœuds N_1, N_2 et N_3, chacune de ces files étant munie d'un ordonnanceur dit de niveau flux : respectivement FLA1, FLA2, FLA3, pour arbitrer entre les flux F1, ..., FN destinés à une même sortie du nœud 1';
- un dispositif ordonnanceur dit de niveau nœud, 2', qui arbitre entre les charges correspondant respectivement aux différentes destinations, en fonction des capacités des bus ;
- quatre dispositifs ordonnanceurs dit de niveau ressource, RA1, RA2, RA3 et RA4.permettant de tenir compte de la façon dont les nœuds N1, ..., N4 sont connectés sur les ressources OR_1, OR_2, OR_3 et OR_4.

La figure 3 illustre cet arbitrage à trois niveaux mis en œuvre dans le dispositif ordonannceur du nœud 1' qui est représenté sur la figure 2.

Bien entendu, l'invention n'est pas limitée aux modes de réalisation qui viennent d'être décrits. En particulier, le nombre de niveaux hiérarchiques peut être supérieure à trois.

Notamment l'invention a été décrite dans le cadre d'un réseau de paquets optiques mais peut être généralisée à tout type de système utilisant des ressources partagées de façon asymétrique tel qu'un système informatique comportant une pluralité d'unités de mémoire (files d'attente) connectées à une pluralité de processeurs (serveurs) via une pluralité de ressources (circuits électroniques) organisées en bus de lecture et d'écriture, le nœud source désignant un composant élémentaire comportant cette pluralité d'unités de mémoire.

De même, les mécanismes d'ordonnancement peuvent être différents de ceux décrits.

REVENDICATIONS

1. Dispositif ordonnanceur (2) pour ordonnancer la transmission de données à partir d'une pluralité de files d'attente (B_1, B_2, B_3) d'un nœud source (1) vers une pluralité de nœuds de destination (N_1, N_2, N_3) via une pluralité de ports de sortie (P_1, P_2, P_3, P_4) dudit nœud source (1), chacun desdits ports de sortie (P_1, P_2, P_3, P_4) étant associés à une ressource (OR_1, OR_2, OR_3, OR_4), les données étant transmises via ladite ressource vers un nœud de destination (N_1, N_2, N_3), chacun desdits nœuds recevant des données de tout ou partie de ladite pluralité de ressources (OR_1, OR_2, OR_3, OR_4), ledit dispositif ordonnanceur (2) étant caractérisé en ce qu'il comporte une pluralité de serveurs (S_1, S_2, S_3, S_4), chacun desdits serveurs étant associé respectivement à une ressource de ladite pluralité de ressources (OR_1, OR_2, OR_3, OR_4) et chacun desdits serveurs comportant des moyens d'ordonnancement, lesdits moyens d'ordonnancement étant indépendants pour chacun desdits serveurs.

2. Dispositif ordonnanceur (2) selon la revendication 1 caractérisé en ce que lesdits moyens d'ordonnancement comportent une pluralité d'étages (L1, L2, L3) correspondant respectivement à une pluralité d'ordonnancements selon des critéres distincts.

3. Dispositif ordonnanceur (2) selon la revendication 1 caractérisé en ce que lesdits moyens d'ordonnancement comportent des moyens d'ordonnancement cyclique du type Round Robin.

4. Dispositif ordonnanceur (2) selon la revendication 1 caractérisé en ce que lesdits moyens d'ordonnancement comportent des moyens d'ordonnancement WFR (Weighted Fair Queueing).

5. Dispositif ordonnanceur (2) selon la revendication 1, caractérisé en ce que lesdits moyens d'ordonnancement sont dépendants d'un ensemble de pondérations statiques et/ou dynamiques.

6. Dispositif ordonnanceur (2) selon la revendication 1, caractérisé en ce que lesdits moyens d'ordonnancement sont dépendants d'un premier ensemble de pondérations, chacune desdites pondérations traduisant le pourcentage d'allocation de ladite ressource à chacun desdits nœuds de ladite pluralité de nœuds.

7. Dispositif ordonnanceur (2) selon la revendication 5, caractérisé en ce que lesdits moyens d'ordonnancement sont dépendants d'un deuxième ensemble de pondérations, chacune desdites pondérations traduisant le poids relatif du trafic de chacun desdits nœuds par rapport au trafic total de la pluralité desdits nœuds.

8. Nœud (1) incluant un dispositif ordonnanceur (2) selon la revendication 1, comportant une pluralité de files d'attente (B_1, B_2, B_3) pour l'émission de données vers une pluralité de nœuds

de destination (N_1, N_2, N_3), et une pluralité de ports de sortie (P_1, P_2, P_3, P_4).

9. Système (10) de transmission de données comportant au moins un nœud source (1) selon l'une des revendications précédentes.